LIDAR-BASED GAIT ANALYSIS IN PEOPLE TRACKING AND 4D VISUALIZATION

Csaba Benedek, Balázs Nagy, Bence Gálai and Zsolt Jankó

Institute for Computer Science and Control, H-1111 Budapest, Kende u. 13-17, Hungary
E-mail: `lastname.firstname@sztaki.mta.hu`

ABSTRACT

In this paper we introduce a new approach on gait analysis based on data streams of a Rotating Multi Beam (RMB) Lidar sensor. The gait descriptors for training and recognition are observed and extracted in realistic outdoor surveillance scenarios, where multiple pedestrians walk concurrently in the field of interest, while occlusions or background noise may affects the observation. The proposed algorithms are embedded into an integrated 4D vision and visualization system. Gait features are exploited in two different components of the workflow. First, in the tracking step the collected characteristic gait parameters support as *biometric descriptors* the re-identification of people, who temporarily leave the field of interest, and re-appear later. Second, in the visualization module, we display moving avatar models which follow in real time the trajectories of the observed pedestrians with synchronized leg movements. The proposed approach is experimentally demonstrated in eight multi-target scenes.

Index Terms— Lidar, gait recognition, 4D reconstruction

1. INTRODUCTION

Efforts on analyzing dynamic 3D (i.e. 4D) scenarios with multiple moving pedestrians receive great interest in various application fields, such as intelligent surveillance, video communication and augmented reality. A critical issue is the assignment of broken trajectory segments during the person tracking process, which problem can be caused by frequent occlusions between the people in the scene, or simply by the fact that the pedestrians may temporarily leave the Field of View (FoV) and re-appear later. People re-identification [1] needs the extraction of biometric descriptors, which may be *weak features* in the sense that we should only focus here on a relatively small group of people, instead of identifying someone from a large database. On the other hand, the targets are non-cooperative and they have to be recognized during their natural behavior.

Gait has been extensively examined as a biometric feature in the recent decades [2]. A video-based gait recognition module may be integrated into surveillance systems in a straightforward way, since it does not need additional instrumentation, and it does not require that the people contact with a special equipment: they may naturally walk distant from the cameras. Although several studies on gait based person identification have been published in literature, most existing techniques have been validated in strongly controlled environments, where the gait prints of the test subjects have been independently recorded one after another, and the assignment has been conducted as an offline process. On the other hand, in a surveillance scenario, the gait features should be observed in an arbitrary scene, where multiple pedestrians are concurrently present in the field, and they may partially occlude each other. To preserve the online analyzing capabilities of the system, the person assignment should also be performed during the action.

A key difficulty with optical camera-based solutions is that view-invariant gait features should be extracted, since it cannot be ensured usually that the subjects are always visible from side view [3]. Multi-camera systems may be utilized to incorporate 3D information in gait analysis [4], however their quick temporary installation is difficult for applications monitoring customized events. Velodyne's Rotating Multi-Beam (RMB) Lidar system is able to provide point cloud sequences from large outdoor scenes with a frame-rate of 15 Hz, 360° FoV, point cloud sizes of around 65K points/frame with 120m radius. The RMB Lidar sensor does not need any installation or calibration process after being placed into a new environment. However, the spatial density of the point cloud is quite sparse, showing a significant drop in the sampling density at larger distances from the sensor, and we can also see a ring pattern with points in the same ring much closer to each other than points of different rings.

Further challenges are related to the efficient visualization of spatio-temporal measurement sequences. Obtaining realistic 4D video flows of real world scenarios may result in a significantly improved visual experience for the observer compared to watching conventional video streams, since a reconstructed 4D scene can be viewed and analyzed from an arbitrary viewpoint, and virtually modified by the user. The *integrated*4D (i4D) vision and visualization system proposed in [5] offers a semi-online reconstruction framework for dynamic 3D scenarios by integrating two different types of data: outdoor 4D point cloud sequences recorded online by a rotating multi-beam (RMB) Lidar sensor, and 4D models of moving actors obtained in an indoor 4D Reconstruction Studio [6]
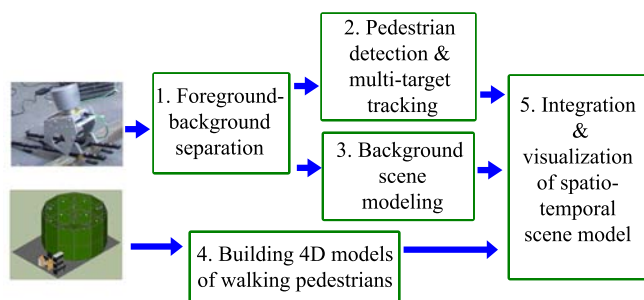
Fig. 1. Workflow of the i4D system

Fig. 2. Silhouette projection: (a) a tracked person and its projection plane in the point cloud from bird's view (b) projection plane from top view, taken as the tangent of the smoothed person trajectory.

in advance (offline). The system is able to automatically detect and track multiple moving pedestrians in the field of interest, and it provides as output a geometrically reconstructed and textured scene with moving 4D studio avatars, which follow in real time the trajectories of the observed pedestrians.

The basic i4D system [5] has two main limitations. *First* only the short time tracking process is implemented, therefore after loosing the trajectories the re-appearing people are always marked as new persons. Although some Lidar based personal identifiers have been adopted in [7] featuring the measured height and the intensity histogram of the people's point cloud segments, experiences showed that these descriptors may confuse the targets, if their heights and clothes are similar. The *second* limitation is connected to the visualization module: although the avatars follow the real person trajectories, turning always according to the trajectory tangent, the animated leg movements are not synchronized with the real walk cycles, causing some visual depreciation. Therefore, gait analysis may also contribute to the improvement of the existing animation framework, so that the actual gait phases are extracted from the Lidar measurements at each time frame, and the phase information is exploited in realistic walking model animation.

In this paper, we aim at overcoming the above mentioned limitations, by supporting the re-identification and animation steps of the i4D system with gait-based features. We will demonstrate that despite the sparseness and inhomogeneity of the provided point clouds, efficient gait analysis can be realized using RMB Lidar sensors, which configuration offers various advantages versus earlier camera based solutions.

2. LIDAR BASED GAIT ANALYSIS

The integrated 4D (i4D) system [5] is a pilot system for reconstruction and visualization of complex spatiotemporal scenes, using a RMB Lidar sensor and a 4D Reconstruction Studio. As shown in the workflow of Fig. 1, the Lidar monitors the scene from a fixed position and provides a dynamic point cloud. This information is processed to build a 3D model of
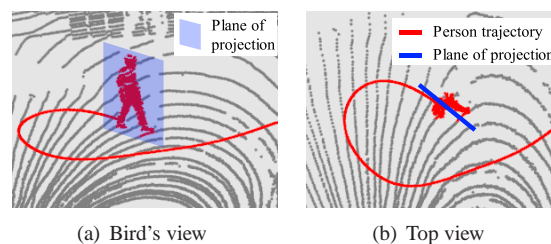
the environment and detect and track the pedestrians. Each of them is represented by a point cluster and a trajectory. A moving cluster is then substituted by a detailed 4D model created in the studio. The output is a geometrically reconstructed and textured scene with avatars that follow in real time the trajectories of the pedestrians.

In the paper we focus on point cloud based gait analysis, supporting the i4D system with two new components:

- In the tracking module (box 2 in Fig. 1), we aim at completing implement a long-term re-identification function with additional gait-based biometric features,

- In the 4D scene visualization module (box 5 in Fig. 1) we intend to synchronize the steps of the moving pedestrians measured in the point cloud sequence, and the leg movements of the animated walking studio objects.

For gait investigation, we follow a 2D silhouette based approach. Several earlier methods rely on the analysis of the *measured* or *interpolated* side view projections of the 3D human silhouettes. Using the considered RMB Lidar point cloud sequences, the side view silhouettes can be estimated in a straightforward way (see Fig. 2). Henceforward, we use the assumption that the people walk forwards in the scene, always turning towards the tangent direction of the trajectory. At each time frame, we project the point cloud segment of each person to the plane, which intersects the actual ground position, it is perpendicular to the local ground plane, and it is parallel to the local tangent vector of the Fourier-smoothed trajectory from top view (see Fig. 2(a),(b)). The projected point cloud consists of a number of separated points in the image plane, which can be transformed into a connected 2D foreground region by morphological operations (Fig. 4(b)).

However, the silhouette extraction step is affected by several practical limitations, as demonstrated in Fig. 4(b). *First*, the point sets of people located far away from the sensor contain less points, therefore the silhouettes may have discontinuities (see Person 2 and Person 4 in Fig. 4(b).) *Second*, for people walking in the direction of the sensor, the 2.5D measurement provides a frontal or back view, where the legs may be partially occluded by each other (see Person 5). *Third*,

some silhouette parts may be occluded by other people or field objects in a realistic surveillance scene.

In our proposed model, we adopted the idea of Gait Energy Image (GEI) based person recognition to the Lidar surveillance environment. The original GEI approach was introduced by Hahn and Bhanu [8] for conventional optical video sequences, and GEIs were derived by averaging the normalized binary person silhouettes over the gait cycles:

$$G(x,y) = \frac{1}{n} \sum_{t=1}^{N} B_t(x,y),$$

here $B_t(x,y) \in \{0,1\}$ is the (binary) silhouette value of pixel (x,y) on time frame t, and $G(x,y) \in [0,1]$ is the (continuous) GEI value. Thus in [8] a person was represented by a set of different GEI images corresponding to the different observed gait cycles, which were compressed by Principal Component Analysis (PCA) and Multiple Discriminant Analysis (MDA). Therafter person recognition was achieved by comparing the gallery (training) and probe (test) features.

In our environment, a number of key differences had to be implemented compared to the reference model [8], leading to a new concept called *Lidar-based Gait Energy Image* (LGEI) representation. First, since the RMB Lidar measurement sequences have a significantly lower temporal resolution (15 fps), than the standard video flows ($\geq$25 fps), samples from a single gait cycle provide too sparse information. For this reason, we do not separate the individual gait cycles before gait print generation, but we select $k = 100$ random *seed frames* from each person's recorded observation sequence, and for each *seed* we average the 60 consecutive frames to obtain a given LGEI sample. In this way, k LGEIs are generated for each individual, which are compressed using a global PCA transform calculated for the whole dataset. Thereafter, the 100 dominant PCA components of the LGEIs are used to train a Multi Layer Perceptron (MLP) for each person. During the training, the LGEIs of the selected person are used as positive samples, while the LGEIs of the others as negative ones. The output of each MLP is a real number $f \in [-1,1]$, where higher f values correspond to high quality matches. In the recognition phase the LGEIs obtained from the test sequences are matched to the MLP model of each trained person, and the decision is taken by comparing the outputs of the different classifiers.

The above sample training and test collection scheme is fully automated and it proved to be notably robust, since in contrast to [8] it is not effected by prior gait cycle estimation. The appearing low quality silhouette frames can be considered as noise in the training data.

The *second part* of the proposed workflow is the synchronization of the observed and animated leg movements in the visualization module. This step needs indeed an approximation of the gait cycles from the Lidar measurement sequence, however the accuracy is not critical here, only the viewer has
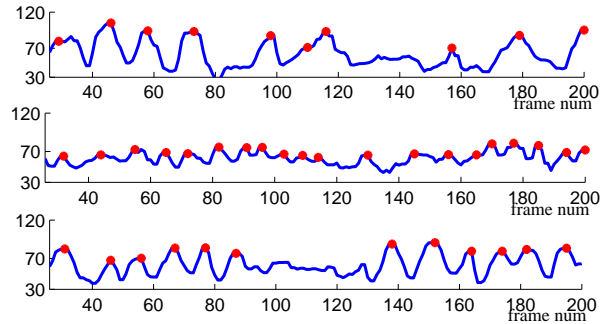

Fig. 3. Silhouette width sequences for three selected persons from the `Walk patterns/1` sequence - used for gait step synchronization during visualization

to be visually convinced that the leg movements are correct. The cycle estimation is implemented by examining the time sequence of the 2D bounding boxes, so that the box is only fitted to the lower third segment of the silhouette. After a median filter based noise reduction, the local maxima of the bounding box width sequence are extracted, and the gait phases between the key frames are interpolated during the animation. Although as shown in Fig. 3, the width sequences are often notably noisy, we experienced that the synthesized videos provide realistic walk dynamics for the users.

3. EXPERIMENTS

We have tested the proposed method in *eight* outdoor sequences captured in a courtyard by a Velodyne HDL 64-E RMB Lidar sensor (sequence names are in Fig. 6). All the sequences have 15 fps frame rate, their length varies between 2:20 and 3:20 minutes, and each one contains 5-7 people walking simultaneously in the scene. In all videos, the test subjects circulate in the scene, then they leave the FoV for a while, and re-appear later in a different order. The goal is to match the corresponding gait patterns collected in the first and second parts of each test scenario.

In the first *three* sequences called `Walk patterns`, the test subjects were asked to follow different characteristic motion types, such as normal walk, jogging or paddling; while in the remaining test walks, called `Spring/1`, `Spring/2`, `Summer/1`, `Summer/2` and `Winter` the subjects moved naturally, therefore we were able to test the separating abilities of the gait descriptors in realistic surveillance situations. Since the sequences were recorded in different seasons, we could also investigate the robustness of the descriptor against the effects on different clothing styles (such as winter coats or t-shirts) on the relevance of the observed gait feature.

Next we provide quantitative evaluation results of people assignment based on LGEIs. For each test sequence, we have trained a MLP for each test subjects in the initial part of the

Observation in realistic surveillance environment:
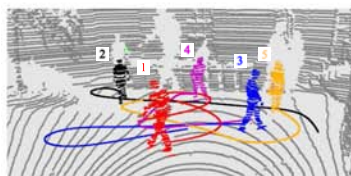
(a) Output of the multi-pedestrian trakcer for a sample Lidar frame (person point clouds+trajectories)

Person 1 ■ Person 2 ■ Person 3 ■ Person 4 ■ Person 5 ■
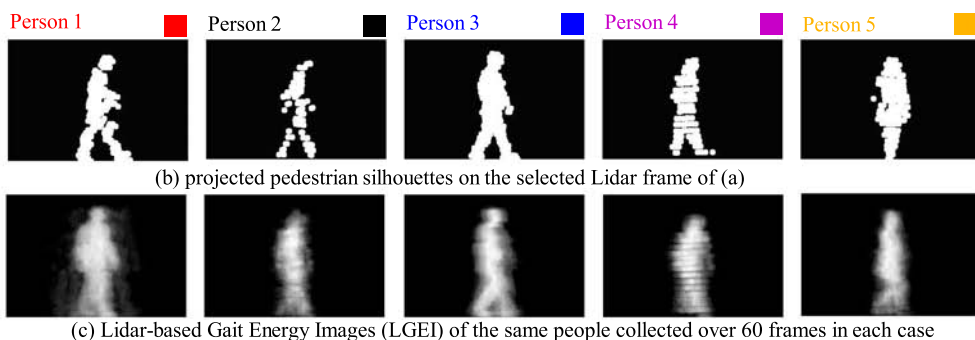
(b) projected pedestrian silhouettes on the selected Lidar frame of (a)

(c) Lidar-based Gait Energy Images (LGEI) of the same people collected over 60 frames in each case

Fig. 4. Demonstration of the (a) 3D multi-target scenario (b) projected silhouettes in a single frame (c) gait prints extracted over a sequence part (60 frames)

scenario. After the re-appearance of the people, we collected LGEI samples for re-identification and calculated the MLP matching scores $f_{ij} \in [-1, 1]$ between the ith trained MLP and the jth extracted test LGEI. The obtained $F = \{f_{ij}\}$ values are displayed in the assignment matrices of Fig. 6, where the rows of the matrices are sorted so that the optimal match is expected in the main diagonal.

We have analyzed the observed results of Fig. 6 at three stages. *First*, considering the F matrix, we have calculated the optimal assignment between the rows and columns with the Hungarian matching [9] algorithm. This process gave perfect assignment in 7 out of the 8 test sequences, while in the Spring/2 test scenario 5 gait patterns were correctly matched, while 2 test subjects were confused, since their available training data were short and had a low quality. Although the above test confirms the efficiency of the LGEI approach, the assignment process requires all training and test samples as input at the same time, which might be not available in online systems. In the *second* test stage, in 6 sequences (except Spring/2 and Summer/2) the optimal matches can also be obtained by a greedy algorithm, i.e. the jth test sample is assigned to the $i^\star$th MLP model where $i^\star = \operatorname{argmax}_i f_{ij}$. If the greedy model works correctly, the recognition can be done immediately after the re-appearance of each test subject, without waiting for the remaining people. In the *third* stage, in 5 out of the 8 sequences the decision of the match between the ith MLP and the jth test sample can be done by simply binarizing the f_{ij} value with a threshold $f_0 = 0.1$. The fulfillment of the later criterion implies the strongest recognition ability, since in this way disappeared or newly appearing people may also be correctly classified in the re-identification phase of the surveillance process. The above tests indicate that the LGEI provides efficient *weak biometric features* for a Lidar based tracker, which can be used together with structural and intensity descriptors [7].

In the visualization module of the i4D system the synchronization of the measurements and the steps of the animated avatars has been implemented. A sample sequence part is
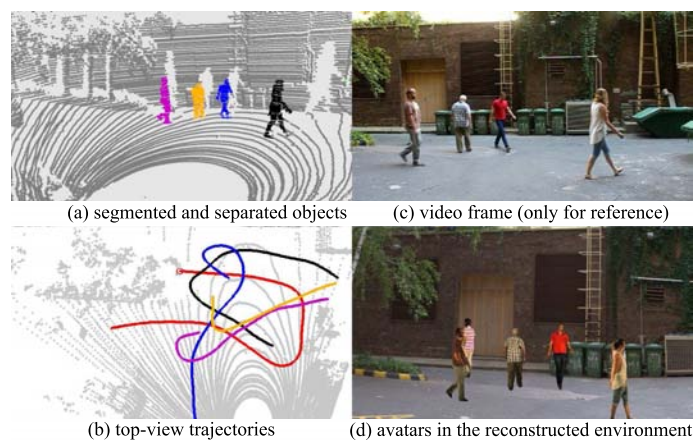

(a) segmented and separated objects (c) video frame (only for reference)

(b) top-view trajectories (d) avatars in the reconstructed environment

Fig. 5. Demonstration of the dynamic scene reconstruction process, with gait step synchronization.

displayed in Fig. 7, showing simultaneously the processed Lidar point clouds, the reference optical video frames and the animated 4D studio object with synchronized steps with the observation. A summarizing figure of the complete recognition and visualization process is displayed in Fig. 5, and the Reader may find demonstrating videos at the following website: http://vimeo.com/user32136096/videos.

4. CONCLUSION

In this paper, we proposed algorithms for gait analysis based on the measurements of a RMB Lidar sensor, in realistic outdoor environments with the presence of multiple and partially occluded walking pedestrians. We have exploited the gait features for person re-identification and 4D visualization tasks. We provided quantitative evaluation results in eight different measurement sequences, and demonstrated the advantages of the approach in future possible 4D surveillance systems.
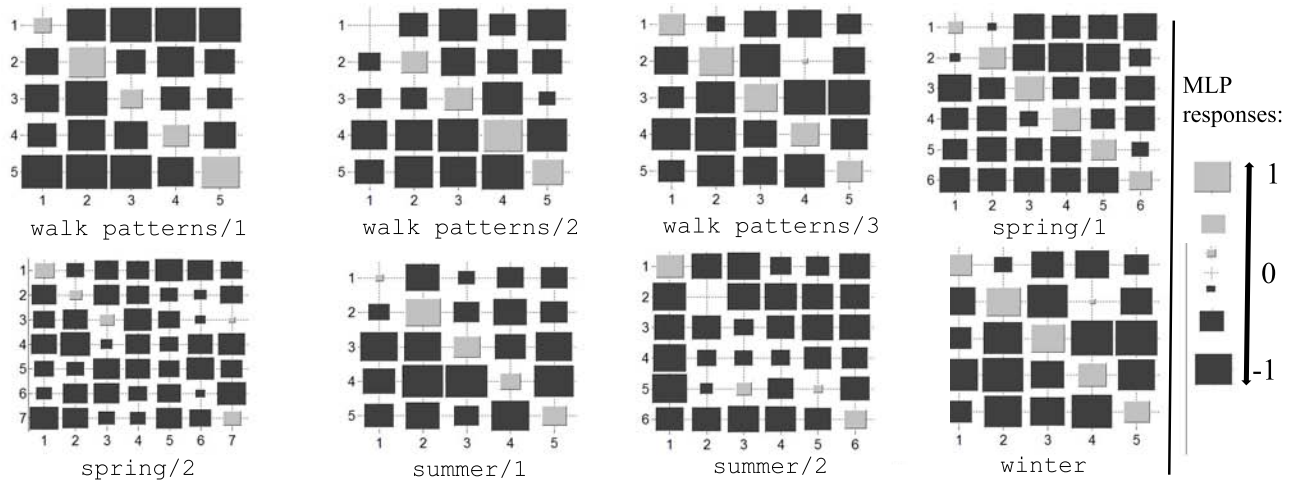
Fig. 6. Quantitative evaluation of LGEI based matching between the gallery (columns) and probe (rows) samples. Rectangles demonstrate the MLP outputs, the ground truth match is displayed in the main diagonal.

(a) Point cloud sequence (used for recognition)

(b) Video images sequence (not used, only visual reference)

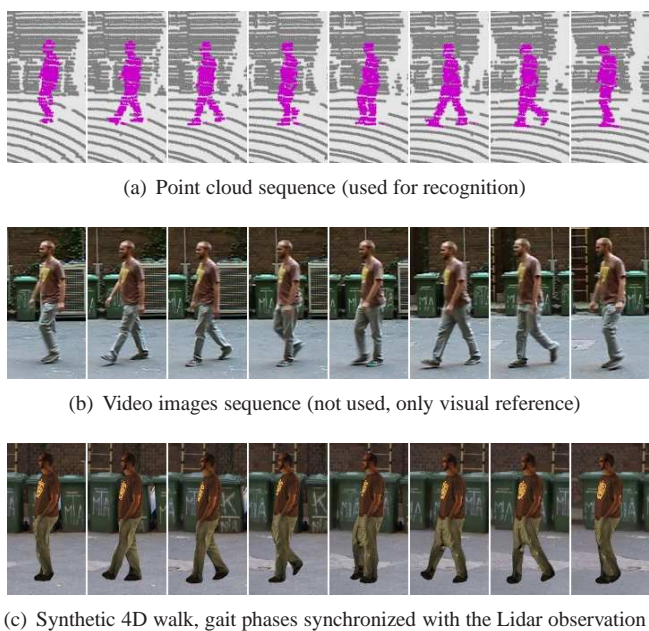(c) Synthetic 4D walk, gait phases synchronized with the Lidar observation

Fig. 7. Sample consecutive frames from the recorded (a) Lidar and (b) video sequences, and the synthesized 4D scene with leg movements synchronized with the observation

REFERENCES

[1] D. Baltieri, R. Vezzani, R. Cucchiara, Á. Utasi, C. Benedek, and T. Szirányi, "Multi-view people surveillance using 3D information," in *Proc. International Workshop on Visual Surveillance at ICCV*, Barcelona, Spain, November 2011, pp. 1817–1824.

[2] Z. Zhang, M. Hu, and Y. Wang, "A survey of advances in biometric gait recognition," in *Biometric Recognition*, vol. 7098 of *Springer LNCS*, pp. 150–158. 2011.

[3] Y. Li, Y. Yin, L. Liu, S. Pang, and Q. Yu, "Semi-supervised gait recognition based on self-training," in *International Conf. Advanced Video and Signal-Based Surveillance (AVSS)*, Beijing, China, Sept 2012, pp. 288–293.

[4] R. Bodor, A. Drenner, D. Fehr, O. Masoud, and N. Papanikolopoulos, "View-independent human motion classification using image-based reconstruction," *Image Vision Comput.*, vol. 27, no. 8, pp. 1194–1206, July 2009.

[5] C. Benedek, Z. Jankó, C. Horváth, D. Molnár, D. Chetverikov, and T. Szirányi, "An integrated 4D vision and visualisation system," in *International Conference on Computer Vision Systems (ICVS)*, vol. 7963 of *Springer LNCS*, pp. 21–30. 2013.

[6] J. Hapák, Z. Jankó, and D. Chetverikov, "Real-time 4D reconstruction of human motion," in *Proc. 7th International Conference on Articulated Motion and Deformable Objects (AMDO 2012)*, 2012, vol. 7378 of *Springer LNCS*, pp. 250–259.

[7] C. Benedek, "3D people surveillance on range data sequences of a rotating Lidar," *Pattern Recognition Letters*, vol. 50, pp. 149–158, 2014, Special Issue on Depth Image Analysis.

[8] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 316–322, Feb 2006.

[9] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistic Quarterly*, vol. 2, pp. 83–97, 1955.